Is Cereal Killer?

Sophia Waxman
Biomedical Engineering
University of California, Davis
Davis, CA
0009-0002-6320-9963

***Abstract*—Breakfast cereals are a common breakfast food in modern diets, yet concerns persist about their nutritional value, particularly regarding sugar content and healthiness. Despite these concerns, the cereal industry remains a multi-billion-dollar market, with many products heavily marketed toward children. This study investigates the nutritional value and consumer ratings of 80 commercially available breakfast cereals from seven major manufacturers, using the publicly available "80 Cereals" dataset. Descriptive and inferential statistical analyses were conducted to explore nutrient composition, consumer ratings, and correlations between cereal characteristics. A composite health score was developed by standardizing key nutritional metrics. Fiber, protein, vitamins, sugar, and calories were all acounted for and adjusted for serving size, allowing cereals to be ranked from least to most healthy. Manufacturer effects on consumer ratings were assessed using analysis of variance methods, and correlation analyses examined the relationships between sugar, fiber content, and cereal type, and potassium levels respectively. The results from these analyses indicate the manufacturer's identity significantly affects consumer ratings, with Nabisco products receiving the highest average ratings. Health score analysis highlights substantial variability within and across manufacturers. Ultimately the cereals marketed towards children as the most unhealthy. Correlation analyses show non-significant relationships between sugar and fiber content, as well as between cereal type and potassium content, suggesting nutrient composition is largely independent of these characteristics. These findings provide quantitative insight into the nutritional landscape of breakfast cereals and demonstrate the power of brand on consumer perception. This work demonstrates the utility of statistical evaluation in assessing dietary quality and informs both consumer choice and future product development.**

***Index Terms*—health, cereal, statistics, sugar, food**

I. Introduction

Breakfast cereals have long been marketed as convenient, nutritious meal options that promise a quick start to the day. Their popularity, particularly in stems from a combination of aggressive marketing, bright packaging, and the perception that they provide essential nutrients efficiently. However, concerns about the nutritional value of these products have persisted for decades. Many cereals, especially those targeted to children, are disproportionately high in sugar while offering minimal amounts of nutrients such as protein, fiber, and potassium [1]. The Environmental Working Group has highlighted that many kids' cereals contain approximately 40% more sugar than is appropriate for a single meal, let alone breakfast. Additionally, the portion sizes listed on packaging often underestimate actual consumption, masking the true amount of sugar ingested daily [1]. These factors raise questions not only about the health impacts of cereal consumption on children but also on broader public health trends.

Despite these nutritional concerns, breakfast cereals remain a multi-billion dollar, multi-national industry. As of 2025, the worldwide market for cereals is estimated at 42.83 billion USD, with a projected compound annual growth rate of 3.54% [2]. This exponential growth is driven by children's preference for sweet flavors and the manufacturers' ability to create brand loyalty through taste and mascot marketing strategies. Moreover, breakfast cereals offer a convenient and shelf-stable option for busy households. This only reinforces their appeal, particularly in large, busy families with limited time for cooking. Understanding the nutritional composition of these cereals is therefore crucial for consumers seeking healthier options and for anyone who wants to increase their food literacy.

This study uses the Kaggle "80 Cereals" dataset [3], which provides a detailed snapshot of the nutritional content and consumer ratings of several cereal products from major manufacturers. The dataset includes quantitative nutritional metrics from FDA mandated labels, such as calories, sugars, fiber, protein, and potassium, as well as categorical information such as manufacturer, cereal type (hot vs. cold), and shelf placement in stores. Although the dataset was last updated in 2017, it remains a valuable resource to analyze patterns in cereal composition and consumer preferences. By combining these metrics with rating data, it is possible to explore the relationship between nutritional quality and perceived appeal to consumers.

The primary goal of this analysis is to assess whether the cereals presented in this dataset can be classified as "healthier" or "less healthy" based on objective nutritional indicators. Additionally, the study investigates whether the manufacturer influences consumer ratings, providing insights into how brand reputation and formulation may affect consumer perception. Specific research questions guiding this work include:

- Which cereals are least healthy based on nutrient composition, including high sugar density, low protein, low fiber, and low vitamin or mineral content?
- Do consumer ratings vary significantly by manufacturer?
- Are sugar and fiber content inversely correlated across cereals?
- Is there a correlation between cereal type and potassium content?

To address these questions, we've employed both descriptive and inferential statistical techniques. Descriptive statistics summarize the distribution of cereals by manufacturer, type, weight, and nutrient density. Additionally, a composite health score is developed by standardizing and weighting key nutritional features, offering a single metric for comparison [4]. Inferential analyses, specifically an analysis of variance (ANOVA), are used to assess whether manufacturer identity significantly influences consumer ratings. Regression modeling also explores relationships between nutritional composition and cereal characteristics.

Beyond the quantitative analysis, this study discusses its findings within the context of the benefits and limitations of cereal consumption. Research indicates that despite concerns about sugar content, regular cereal consumption may contribute positively to diet quality [5]. This is particularly true of high-fiber or nutrient-fortified options [5]. Cereal can serve as a practical source of whole grains, fiber, and micronutrients, supporting both short-term energy needs and long-term health outcomes. This is particularly true for people who would otherwise skip breakfast due to time constraints. However, careful selection is critical. Cereals with excessive sugar and minimal nutrients can exacerbate health risks rather than mitigate them.

By combining data visualization, descriptive statistics, and statistical modeling, this research looks at patterns in the nutritional landscape of commercial cereals, evaluates consumer perceptions, and identifies which cereals are the healthiest and the least healthy along a variety of metrics. These insights are valuable not only for consumers seeking to optimize their breakfast options but also for nutrition educators, statisticians, and manufacturers interested in developing and promoting their cereal products. Ultimately, the study emphasizes the importance of evidence-based assessment in understanding the complex trade-offs between taste, convenience, and nutritional value in the breakfast cereal market.

II. Methods

A. The Data Set

The dataset used in this study is the "80 Cereals" dataset, published on Kaggle [3]. The dataset was originally compiled in 1993 by the American Statistical Association as part of a teaching dataset collection designed to support data literacy and applied statistics education. Although the version used here was last updated in 2017 when it was uploaded to Kaggle, it continues to provide a representative snapshot of the U.S. cereal market, capturing the nutritional composition and consumer ratings of commercially available breakfast cereals.

The dataset includes cereals produced by seven major manufacturers:

- A: American Home Foods Products
- G: General Mills
- K: Kellogg's
- N: Nabisco
- P: Post
- Q: Quaker Oats
- R: Ralston Purina

Each row in the dataset represents one cereal product, described by a combination of categorical and numerical attributes. The categorical variables include the manufacturer, cereal type, and shelf placement. The continuous variables encompass caloric content, protein, fat, sodium, dietary fiber, carbohydrates, sugars, potassium, vitamin enrichment, serving weight, serving volume, and consumer rating.

Data were imported directly from Kaggle using the RKaggle package in R to ensure reproducibility. All data analysis was performed in RStudio using the tidyverse, ggplot2, GGally, and pwr packages for data cleaning, visualization, and statistical modeling. The use of open-source tools allows full transparency and replicability of the results.

B. Data Processing

All processing and analysis scripts are publicly available on GitHub at:

github.com/30ph1a/MAE298IsCerealKiller.

The dataset contained several missing or invalid entries coded as -1, which were treated as missing values and excluded from any analysis requiring those specific variables. After cleaning the data, there was a total of 77 cereals included in this analysis. Variables were standardized where appropriate with z-scores to ensure comparability across different nutrient scales. Data exploration was conducted using summary tables, histograms, and various statistical tests to identify outliers, data entry inconsistencies, and skewed distributions.

To facilitate reproducibility, all transformations, visualizations, and statistical tests were executed within a single R Markdown file. This ensures transparency and allows any researcher to replicate the study by downloading the GitHub repository and running the R script end-to-end. The repository also includes a pdf output file, which displays the results. It is linked above.

C. Data Analysis

The analysis performed used both descriptive and inferential statistical methods. Descriptive statistics were used to summarize manufacturer distributions, cereal types, and nutrient concentrations. Correlations, linear regressions, and ANOVA were used to test relationships between nutrients, manufacturer identity, and consumer ratings. All data analysis was performed through the R computing suite.

1) Descriptive Statistics: The descriptive analysis provided an overview of the dataset's structure and characteristics. Of the 77 cereals, Kellogg's and General Mills were the most represented manufacturers, while American Home Foods Products had only one entry. This lack of data points made it necessary to remove the manufacturer from the data set for some analysis. The vast majority of cereals were classified as cold cereals, with only 3 products designated as hot cereals, reflecting the dominance of ready-to-eat cold cereals in the U.S. market.

The shelf placement analysis revealed subtle variations in average serving weights. These weights were standardized by serving size. The number of servings in a box were not accounted for, as no data was collected on it. Cereals placed on the top shelf had the heaviest average weight, followed by the bottom, and the middle shelf. This pattern may suggest that products placed higher on shelves and at eye level for adults, tend to be denser, than those at lower, child height eye levels.

The sugar density, defined as grams of sugar per ounce of cereal, was calculated to assess the relative concentration of added sugars. Golden Crisp and Smacks, products of Post and Kellogg's respectively, marketed towards children, had the highest sugar density among all cereals. This confirms findings by the Environmental Working Group that many children's cereals are extremely sugar-dense products [1]. Quaker Oatmeal was removed from this analysis, due to missing values Additionally, fiber content averages across manufacturers revealed that Nabisco produced the most fiber-rich cereals overall, with Post coming second.

2) Which cereals are the least healthy?: The concept of "healthiness" was defined in this study based on a balance of beneficial and detrimental nutrients. A healthy cereal was defined as one that maximizes fiber, protein, and vitamin content while minimizing sugar and caloric density, following established nutritional guidance from Harvard Health [4]. To quantify this, nutrient variables were standardized (z-scores) and combined into a composite health score using the following equation:

$$\text{health} = \frac{f + p + v - (s + c)}{CS} \tag{1}$$

where f, p, and v represent standardized fiber, protein, and vitamin scores, respectively. The variables s and c represent standardized sugar and calorie scores, and CS represents the serving size in cups.

Using this measure, the five cereals determined to be healthiest are characterized by high fiber and vitamin content with low sugar levels. The least healthy cereals exhibited high sugar densities and minimal fiber. These findings align with previous reports emphasizing the nutritional disparity between adult-oriented and child-targeted cereals [1], [4]. The healthy and unhealthy cereals also contained varied representation between manufacturers.

3) Does the rating of the cereal differ by manufacturer?: To assess whether consumer ratings differ significantly by manufacturer, an ANOVA was conducted using the model:

$$\text{rating} \sim manufacturer \tag{2}$$

Because American Home Food Products had only one cereal in the dataset, it was removed from this analysis to prevent distortion of the variance estimates. It has remained in the data visualization, however. The initial model indicated a statistically significant effect of manufacturer on cereal ratings, ($F(6, 70) = 6.804$, $p = 1.03 \times 10^{-7}$), suggesting that mean consumer ratings differ meaningfully between brands, though the ANOVA does not identify exactly which brand or brands differ.

Following this finding, post-hoc Tukey HSD pairwise comparisons were conducted to determine which manufacturers' ratings differed significantly from one another. These results revealed that Nabisco cereals had consistently higher ratings than cereals produced by all other manufacturers, with significant differences observed at the $\alpha = 0.001$ level or lower. This indicates that Nabisco's products were perceived by consumers as notably superior in the rating scheme used, when compared to competing brands.

During model diagnostics, visual inspection of residuals and influence measures revealed a potential outlier. Cook's distance analysis identified All-Bran with Extra Fiber, a Kellogg's brand cereal, as a high-influence data point due to its exceptionally high consumer rating of 93.7. To test the robustness of the model, this outlier was removed and the ANOVA was re-run. After this adjustment, the model remained highly significant, with ($F(6, 69) = 9.114$, $p = 2.45 \times 10^{-7}$). This suggests that the significance discovered is not reliant on the outlier and there is a true significance in who the manufacturer is when determining the rating.

Re-running the Tukey HSD analysis after outlier removal confirmed the earlier pattern: Nabisco cereals continued to receive significantly higher ratings than those of any other manufacturer, with all pairwise comparisons significant at the $\alpha = 0.0005$ level or lower. No other inter-manufacturer comparisons reached statistical significance.

These findings suggest that the manufacturer Nabisco's identity plays a meaningful role in consumer perception. The other manufacturers showed no significance in determining their ratings, with their identities having no bearing on rating. This preference may be due to brand power, a greater grasp of what consumers are looking for, or perhaps other contributing factors.

4) Are sugar and fiber inversely correlated?: To test whether higher sugar content is associated with lower fiber content, a Pearson correlation was computed between sugar and fiber values. The correlation coefficient was $r = -0.139$ ($p = 0.232$), indicating a weak, statistically insignificant inverse relationship. This suggests that while high-sugar cereals tend to have lower fiber, this trend is not consistent enough across brands to be conclusive and we cannot say the difference is not zero. We cannot conclude there is a correlation.

5) Is the type of cereal correlated with potassium content?: Finally, a Pearson correlation test was used to determine whether the type of cereal, hot or cold served cereal, is associated with potassium content. Cream of Wheat and Almond Delight were removed from this analysis due to missing values. The correlation was small and non-significant ($r = 0.080$, $p = 0.490$), suggesting that cereal type does not meaningfully predict potassium levels. Given that only three cereals were categorized as "hot," the statistical power for this test was low, limiting interpretability. As it stands we

cannot conclude there is a correlation. More analysis is needed between hot and cold cereals.

III. Results

A. Manufacturer Distribution and Cereal Types

The dataset contained 80 cereals from seven major manufacturers: American Home Foods, General Mills, Kellogg's, Nabisco, Post, Quaker Oats, and Ralston Purina. Each brand was coded in the data with a single letter indicator. Kellogg's and General Mills produced the largest variety, with 23 and 22 cereals respectively, while American Home Foods had only one product. The remaining manufacturers offered between six and nine cereals each, indicating a highly concentrated market dominated by a few major brands.

When categorized by cereal type, the overwhelming majority were cold cereals (74 of the 77 total data points, after excluding missing values), while only three data points were classified as hot cereals. This confirms that cold cereals dominates the commercial breakfast market represented by this dataset.

B. Shelf Placement and Product Characteristics

To explore potential marketing or packaging trends, average weight was compared across shelf levels, accounting for serving size differences. Cereals placed on shelf 1, the lowest shelf, averaged 1.160 oz/cup, shelf 2 averaged 1.159 oz/cup, and shelf 3 averaged 1.684 oz/cup. This modest trend suggests that cereals displayed on the highest shelf tend to be slightly heavier on average than other shelves, which weighed approximately the same. The difference recorded is small and may not be practically significant.

C. Nutritional Content Comparisons

To evaluate differences in nutrient composition, sugar density, defined as grams of sugar per ounce of cereal by serving size, was also calculated. Golden Crisp and Smacks achieved the highest score of sugar concentration, tied at 15 g/oz. This indicates that they are the most sugar-dense cereal.

Fiber content was compared across manufacturers to determine which company produced the most fiber-rich products. Nabisco, with an average of 4 g had the highest average fiber content among all manufacturers, suggesting a stronger nutritional focus on fiber within its product line.

D. Healthiness Scores

To quantify overall nutritional quality, a health score was computed for each cereal according to the standardized formula presented in Equation 1. Cereals with the highest health scores included All-Bran, a Kellogg's product, 100% Bran from Nabisco, All-Bran from Kellogg's, Grape-Nuts from Post, and Total Whole Grain from General Mills. These products scored highly due to elevated fiber and protein levels and low sugar and calorie contents, as described in the equation.

Conversely, the lowest health scores were found for Frosted Flakes a Kellogg's product, Cinnamon Toast Crunch from General Mills, Fruity Pebbles from Post, Mueslix Crunch from Kellogg's, and Cap'n Crunch from Quaker Oats. These cereals had high sugar content and low fiber and protein levels, aligning with their reputation as less nutritious options.

E. Differences in Ratings by Manufacturer

A one-way analysis of variance (ANOVA) was conducted to assess whether consumer ratings differed significantly across manufacturers. American Home Foods was excluded from this test due to having only one observation. The initial model showed a significant main effect of manufacturer, $F(6, 70) = 6.804$, $p = 1.03 \times 10^{-5}$, indicating that mean ratings varied by brand. This test did not determine which brand or brands differed and futher testing was required.

Tukey HSD post-hoc comparisons revealed that Nabisco cereals received significantly higher ratings than those produced by the other recorded manufacturers, while differences among other manufacturers were not statistically significant. This result was confirmed through visual inspection with Figure 1.

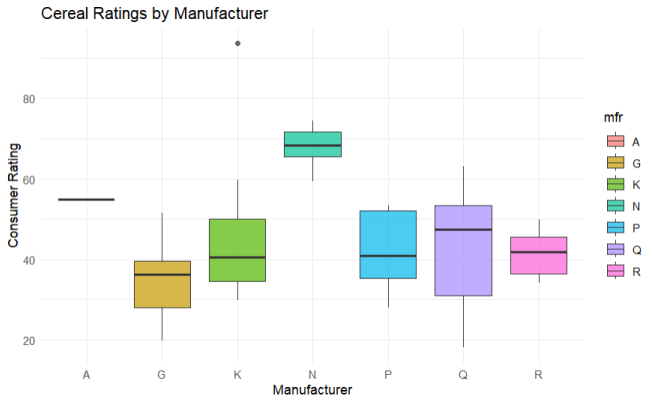

Fig. 1. Distribution of Cereal Ratings by Manufacturer.

Additionally, visual inspection of Figure 1 and Cook's distance diagnostics identified All-Bran with Extra Fiber, a Kellogg's product, as a highly influential outlier due to its exceptionally high consumer rating (93.7). The Cook's distance is shown in Figure 2, with the outlier denoted with a red asterisk.

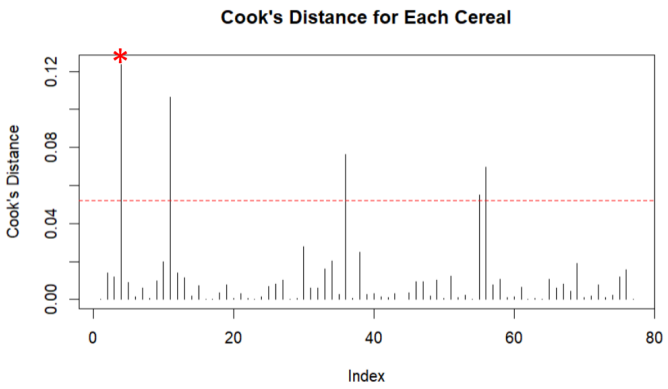

Fig. 2. Cook's distance of the data points with the outlier (All-Bran with Extra Fiber marked with a red asterisk.

As the goal of this analysis is to study the average differences between manufacturers and the data point is highly influential, the data point was removed. Though significance was already proved, looking at the data without the outlier was deemed to be of interest. After removing this single outlier, a second ANOVA was run. The manufacturer effect remained significant and became even stronger, $F(6, 69) = 9.114$, $p = 2.45 \times 10^{-7}$. The results were again visualized and are shown in Figure 3.

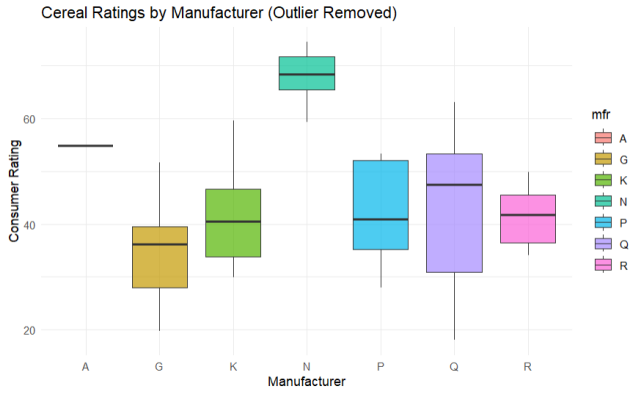

Fig. 3. Cereal Ratings by Manufacturer without the outlier (All Bran with Extra Fiber).

Post-hoc analysis again confirmed that Nabisco cereals consistently outperformed all other brands, maintaining significantly higher consumer ratings at the $\alpha = 0.0005$ level or lower. This suggests that Nabisco products are perceived as more appealing by consumers, even after accounting for data outliers. This result is reliant on the metrics and evaluation criteria of the evaluating body and rigor of the survey.

F. Correlation Analyses

Two Pearson correlation tests were additionally conducted to explore nutrient relationships.

First, the correlation between sugar and fiber content was weakly negative and non-significant ($r = -0.139$, $p = 0.232$), suggesting that cereals with higher sugar content do not necessarily contain less fiber. In fact, the study cannot conclude the correlation is not zero, as the 95% Confidence Interval [-0.35, 0.089] includes zero within its range. Quaker Oatmeal was removed from this analysis due to missing data values.

Second, a correlation test between cereal type and potassium content was also non-significant ($r = 0.080$, $p = 0.490$). This test's 95% Confidence Interval [-0.15, 0.30] also included zero. Given that only three cereals in the dataset were hot cereals, this result should be interpreted cautiously due to limited statistical power. The small correlation implies that potassium levels are likely not influenced by hot or cold cereal type. Cream of Wheat and Almond Delight cereals were removed from this analysis due to missing values. This further decreased the power the statistical analysis was able to obtain, as Cream of Wheat is one of the three hot cereals. This left only two of the hot cereals available for analysis.

IV. Discussion

The results of this study reveal clear patterns in the breakfast cereal market that reflect both manufacturer preferences and broader nutritional trends. The larger representation of Kellogg's and General Mills Cereals within the dataset demonstrates the higher degree of market concentration of these brands. These findings are consistent with broader market research indicating that a few large multinational corporations control most global cereal production and sales [2]. Such concentration not only shapes product availability but also influences consumer perceptions through extensive branding and marketing strategies that reinforce brand loyalty. The breakfast cereals become more than a meal option with this reinforcement. They become an institution and must-have.

The near exclusivity of cold cereals in the dataset further illustrates the market's emphasis on convenience and rapid preparation. These are features valued by consumers with busy lifestyles. This aligns with global market trends showing growing demand for ready-to-eat food options, driven by urbanization and the expansion of middle-income populations [2]. The predominance of cold cereals may also reflect retailers' shelf allocation strategies, which prioritize high-turnover products with wide appeal.

Nutritional comparisons among manufacturers demonstrated notable differences in product formulation strategies. Nabisco's higher average fiber content differs sharply with the elevated sugar levels found in Golden Crisp, Smacks, and similar brands. This variation suggests a clear separation in cereal manufacturers product design. Some are created to appeal to health-conscious adults and some more for younger, sugar-oriented demographics. The Environmental Working Group has previously documented this divide, reporting that many children's cereals contain much more sugar than adult-oriented varieties [1]. Such disparities highlight an enduring division in the cereal market between nutritional content and taste, a challenge frequently discussed in nutrition and public health research [4].

Statistical testing revealed significant differences in cereal ratings by manufacturer, with Nabisco products performing better over all the others. The identification and removal of an influential outlier, All-Bran with Extra Fiber, demonstrated the sensitivity of the statistical analysis to extreme values. Though the analysis already confirmed the significance of the result, Eliminating the outlier only increased the significance.

Correlation analyses yielded limited associations among key nutritional variables. The lack of strong relationships between sugar and fiber, or between cereal type and potassium content, suggests that product formulation is not driven by simple linear trade-offs among nutrients. Manufacturers likely optimize formulations based on a combination of multiple factors and considerations, rather than adhering to a fixed nutrient profile. While breakfast cereals can contribute significantly to dietary fiber, iron, and B-vitamin intake, the health impact of cereal consumption depends heavily on the specific product chosen [5]. This variability underscores the need for nuanced, product-level nutritional assessments rather than broad generalizations about cereal healthfulness. Further data is also needed to confirm these analyses, particularly for the cereal type and potassium analysis. This particular analysis lacked the power for any definitive result with only two usable hot cereal data points.

Taken together, these findings emphasize the complex and sometimes contradictory nature of the cereal market with each cereal seemingly formulated seperatly from other avalible options. The coexistence of highly sugary, child-oriented cereals alongside fiber-rich, health-focused products illustrates how manufacturers balance profitability with public health pressures, particularly for a market split between adults' and children's wants. As consumer awareness of sugar content and nutritional labeling grows, major brands may face increasing pressure to reformulate products or diversify offerings toward healthier alternatives [4]. Future research could expand on these results by incorporating data on product reformulations, sales trends, and evolving consumer preferences, providing a more comprehensive understanding of how nutritional and market dynamics interact over time. Additionally, further data on hot cereals is necessary to better acount for formulation and desirablitity differences.

V. Conclusion

This study analyzed breakfast cereals produced by seven major manufacturers, examining their nutritional composition, product characteristics, and consumer ratings to identify underlying trends in the cereal market. The results provide an integrated view of how industry practices align with consumer preferences and health considerations. It also illustrates what further data is necessary to fully understand the topic of cereal. Key findings include:

- Market concentration among leading manufacturers: Kellogg's and General Mills accounted for the majority of cereals in the dataset, reflecting their dominance in the breakfast cereal market. This concentration mirrors broader industry patterns, where large multinational firms maintain substantial market shares through brand recognition, product diversification, mascot creation, and aggressive marketing strategies [2]. Smaller manufacturers, while contributing fewer products, often target niche markets with health-oriented or specialty cereals and are largely not well represented in this data set.
- Prevalence of cold cereals and implications for consumer behavior: Cold cereals overwhelmingly dominated the dataset, underscoring consumer preferences for convenience, speed, and portability. This trend aligns with global market data showing that ready-to-eat cereals continue to outperform hot cereal segments in both volume and value [2]. The finding also reflects lifestyles and time constraints that favor quick breakfast options, particularly in urbanized societies. People work hard and don't have the time to spend preparing a warm meal, so they utilize cold cereal options.
- Shelf placement and product characteristics: The observed association between the highest shelf placement and slightly heavier products suggests strategic packaging and marketing choices. Positioning heavier or premium cereals at eye level on upper shelves may enhance visibility and perceived quality, as people are most likely to chose an option easily visible to them. While the effect was modest, it underscores the interplay between product design, marketing psychology, and retail strategy.
- Variation in nutritional profiles among manufacturers: Nutritional composition varied substantially across brands. Nabisco produced cereals with the highest average fiber content, aligning with a health-oriented branding strategy. In contrast, Golden Crisp and Smacks both tied to exhibit the highest sugar density, characteristic of cereals targeted toward children and younger consumers. These results echo the Environmental Working Group's findings that many children's cereals contain substantially higher sugar levels than those marketed to adults [1]. This differentiation suggests that manufacturers segment the market based on consumer health awareness and taste preferences.
- Statistical analysis of manufacturer effects and outliers: ANOVA and post-hoc comparisons demonstrated significant differences in cereal ratings between manufacturers, with Nabisco products rated highest overall. The analysis also identified an influential outlier, A Kellogg's product, All-Bran with Extra Fiber, whose removal enhanced the strength of the manufacturer effect. This finding underscores the importance of outlier detection and deciding on keeping/removing them, especially when sample sizes differ among groups.
- Inter-nutrient relationships and formulation complexity: Correlation analyses indicated no strong linear relationships between sugar and fiber content or between cereal type and potassium levels. This suggests that nutritional attributes are not tightly coupled, reflecting the complexity of cereal formulation, where manufacturers balance taste characteristics, nutrient fortification, and production

costs independently. It should be noted that cereals vary widely in nutrient content even within the same product category and that overall dietary contribution depends heavily on specific product choices [5].

Overall, this research demonstrates the value of applying statistical and exploratory data analysis to consumer food products. The findings reveal how manufacturer strategies, nutritional composition, and consumer ratings intersect to shape perceptions of health and quality. Beyond descriptive insights, the results highlight the potential of quantitative and inferential methods, such as ANOVA, correlation analysis, and outlier diagnostics, to clarify relationships within complex consumer datasets.

Future studies could expand on this work by incorporating larger and more diverse datasets, including international cereal brands and newer market entrants, to improve representativeness. Integrating sales, pricing, or reformulation data would provide greater insight into how nutritional attributes affect commercial performance. Moreover, multivariate modeling and machine learning approaches could be used to predict cereal ratings or health scores based on combinations of nutritional and marketing variables, offering a more comprehensive understanding of the factors driving both product quality and consumer preference in the evolving breakfast cereal market. Greater data diversity could also be useful in further analyzing this.

VI. Acknowledgment

I would like to thank Professor Rich S. Whittle for his guidance and instruction in MAE 298: Design of Experiments and Statistical Methods for Engineers. This project benefited greatly from the insights provided in the course, including statistical analysis techniques, data interpretation, and experimental design principles. I'd also like to thank Kaggle and Chris Crawford for compiling the data. In the form I found it in. I'd also like to thank the American Statistical Association Section on Statistical Graphics as the originators of the data set.

References

[1] Environmental Working Group, "Kids' Cereals Pack 40 Percent More Sugar," EWG News & Insights, Washington, D.C., May 15, 2014. Available: https://www.ewg.org/news-insights/news-release/kids-cereals-pack-40-percent-more-sugar?

[2] V. Swar and A. Shivarkar, "Breakfast Cereal Market Size to Hit USD 58.58 Billion by 2034", Precedence Research, Nov. 4, 2025. Available: https://www.precedenceresearch.com/breakfast-cereal-market

[3] Crawford, Chris. "80 Cereals." Www.kaggle.com, 2017, www.kaggle.com/datasets/crawford/80-cereals?resource=download. Accessed 6 Nov. 2025.

[4] Shmerling, Robert H. "Is Your Breakfast Cereal Healthy? - Harvard Health." Harvard Health, 8 Sept. 2025, www.health.harvard.edu/blog/is-your-breakfast-cereal-healthy-202509083104. Accessed 6 Nov. 2025.

[5] Williams, Peter G. "The Benefits of Breakfast Cereal Consumption: A Systematic Review of the Evidence Base." Advances in Nutrition, vol. 5, no. 5, 1 Sept. 2014, pp. 636S673S, https://doi.org/10.3945/an.114.006247.